

Dokumentation des Ansichtskartenkorpus

Vorgehen bei der Digitalisierung¹

1 Einleitung

In diesem Anhang werden die Vorgaben aufgezeigt, nach denen die gescannten Ansichtskarten erfasst wurden. Er gibt einen Einblick in die Vielfalt der erfassten Textkörper- und Metadaten, welche sowohl für unser Forschungsprojekt als auch für künftige Studien zur Verfügung stehen.

Ausgangslage: Es wurden über 12.000 Ansichtskarten abgeschrieben und in einer Online-Eingabemaske erfasst. In der Sammlung sind auch Karten, welche nicht abgeschrieben (aber aufbewahrt) wurden:

- Ansichtskarten, die nicht auf Standarddeutsch oder Schweizerdeutsch verfasst wurden
- Geburtstags- und Weihnachtskarten, die nicht aus dem Urlaub geschrieben wurden
- Karten mit leerem Adressfeld
- Karten ohne Bildseite
- alte Karten, bei denen sich der Mitteilungstext auf der Bildseite befindet (vgl. Abbildung 1a und 1b)

1 Diese Dokumentation ist im Forschungsprojekt »Textsortenentwicklung zwischen Standardisierung und Variation: Das Beispiel der Ansichtskarte. Text- und korpuslinguistische Untersuchungen zur Musterhaftigkeit privater Fern- und Alltagsschriftlichkeit« als Richtlinie für das Schreibbüro entstanden, das die Transkription und Annotation durchgeführt hat. Für diesen Sammelband hat Kyoko Sugisaki die ursprüngliche Richtlinie überarbeitet, die durch die folgenden Projektmitarbeitenden verfasst wurde: Nicolas Wiedmer, Michaela Schnick, Marcel Naef, Kyoko Sugisaki, Dewi Josephine Obert, Maaike Kellenberger, Selena Calleri, David Koch.

Abb. 1a und 1b: Text auf Bildseite



Quelle: Zürcher Ansichtskartenkorpus (gescannt, aber Text nicht digitalisiert)

2 Eingabemaske

Die Karten wurden in eine Eingabemaske (Abbildung 2) abgeschrieben, in der Informationen zu unterschiedlichen Dimensionen des Dokuments erfasst werden konnten. Jede gescannte Karte wurde durch ein Schreibbüro in Deutschland sortiert, abgeschrieben und erfasst. Die Maske wurde speziell fürs Schreibbüro entwickelt, damit eine große Menge an Karten (über 14.000 Karten als Ausgangslage) schnell, effizient und möglichst fehlerfrei ohne technisches Know-how, z. B. im Bereich XML-Technologie, verarbeitet werden konnte. Speziell die folgenden Anforderungen stellten sich als Herausforderung heraus: zwei Bilder (Bild- und Schriftseite) dienen als Vorlage, und sowohl Metadaten als auch die Textannotation mussten pro Objekt erfasst werden, inkl. automatischer formaler Überprüfung. Die bestehenden Annotationstools, zumindest zum Stand des Projekts im Jahre 2016, waren solchen besonderen Use-Cases nicht gewachsen. Technische Details der Maske sind in Sugisaki et al. 2018 beschrieben.

Abb. 2: Eingabemaske



Quelle: <https://postcards.linguistik.uzh.ch>

3 Annotation

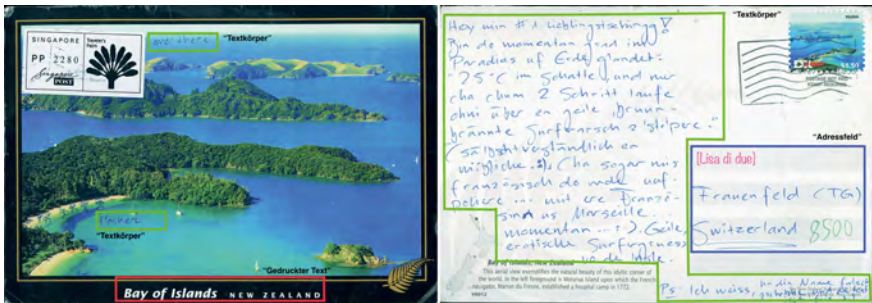
- Für die Transkription und die Annotation der Karten wurden Richtlinien fürs Schreibbüro erarbeitet. Die folgenden Ausführungen geben Einblick in die Richtlinien und die dahinterstehenden Entscheidungen darüber, wie die einzelnen Karten abgeschrieben und erfasst werden sollten, um dem gegenwärtigen und zukünftigen Forschungsvorhaben zu genügen.

3.1 Definitionen der Karte

Auf den Karten werden verschiedene Elemente unterschieden:

- *Schriftseite* (vgl. Abbildung 3a):
 - o Textkörper (gesamter handschriftlicher Text außerhalb des Adressfelds)
 - o Adressfeld
- *Bildseite* (vgl. Abbildung 3b)
 - o Textkörper (gesamter handschriftlicher Text)
 - o Gedruckter Text

Abb. 3a und 3b: Schrift- und Bildseite



Quelle: [anko] 70123

3.2 Transkription und Annotation des Textkörpers

Wir nennen den gesamten handschriftlichen Text außerhalb des Adressfelds »Textkörper«. Der Textkörper auf der Schrift- und der Bildseite wird während des Abschreibens nicht nur erfasst, sondern auch annotiert. Annotation bedeutet die Auszeichnung bestimmter Passagen des Textkörpers wie Anrede oder Datum mittels spezieller Zeichenfolgen. Diese werden »Tags« genannt.

Es gibt zwei verschiedene Arten von Tags:

- Tags mit eckigen Klammern, welche anstelle der Einheit gesetzt werden, die sie markieren – *Beispiel*: Persönliche Informationen, wie eine Telefonnummer, werden durch [vertraulich] ersetzt.
- Tags als Kombinationen von senkrechtem Strich und Sternchen sind als öffnende und schließende Markierung eingesetzt. Öffnende und schließende Tags sind spiegelverkehrt, und die Anzahl der Sternchen ist ein differenzierendes Merkmal für verschiedene Annotationstypen – *Beispiel*: Der Tag, welcher die Anrede markiert, wird umschließend gesetzt: **|Hallo Anna|**

Tags können von anderen Tags umschlossen sein. Zwei direkt aufeinander folgende Tags werden durch ein Leerzeichen voneinander getrennt. Im folgenden Beispiel wird der Tag [NN] als Ersetzung für einen Nachnamen (vgl. Abschnitt Nachname) von einem Tag für »Anrede« umschlossen:

- *Beispiel*: wenn in der Anrede *Liebe Frau Meier* steht, wird das als **|Liebe Frau [NN] |** getaggt.

Weitere Vorgaben:

- Der Textkörper soll exakt (zeichengetreu) abgeschrieben werden. Regionale Dialekte und Rechtschreib-/Grammatikfehler sind 1:1 zu übernehmen und sollen nicht korrigiert werden.
- Wenn etwas vom Verfasser korrigiert wurde, wird nur die korrigierte Version abgeschrieben.
- Zeichnungen wie Piktogramme oder Ideogramme, die Teil des Textes sind, werden so abgeschrieben, wie man Sie vorlesen würde: z.B. wird *sung!* als *Achtung!* und ♥ *Grüsse* als *Herzliche Grüsse* abgeschrieben. Abtippbare Emoticons werden in gleicher Form wie auf der Karte erfasst. Zusätzlich sind Zeichnungen und Emoticons jeglicher Art in den Metadaten bei den Punkten »Verzierung Schriftseite« und/oder »Verzierung Bildseite« als »vorhanden« anzugeben.
- Wenn auf der Schrift- oder auf der Bildseite kein handschriftlicher Text vorhanden ist, wird ein *x* eingegeben.

Für die Annotation des Textkörpers gelten die folgenden Vorgaben:

Datum

- *|Datum|*
- Beispiele: *|7. 3. 85|* oder *|August 1997|*
- Es werden nur handschriftliche Datumsangaben im Textkörper annotiert und nicht das Datum des Poststempels.
- Das Datum wird zeichengetreu abgeschrieben.

Anrede und Grußfloskeln

- *Anrede*
 - **|Anrede|**
 - Beispiele: **|Lieber Daniel!|** oder **|Hallo zusammen,|**
 - Die Anrede steht meistens am Anfang der Mitteilung.
 - Steht nach der Anrede ein Satzzeichen, wird es als Teil der Anrede annotiert.
- *Gruß*
 - ***|Gruß|***
 - Beispiele: ***|Liebe Grüsse aus Mexico!|*** oder ***|Mit Liebe, Deine|*** oder ***|Bis bald, dein dich liebender|***
 - Grüße stehen meistens unmittelbar vor den Unterschriften. Sie können aber auch am Anfang oder mitten in der Mitteilung stehen.
 - Auch wenn der handschriftliche Text nur aus einem Gruß besteht, wird er trotzdem als Gruß getaggt.
 - Zusätze wie *Dein, Eure, von* etc. gehören zur Grußformel.
 - Folgt auf den Gruß ein Satzzeichen, wird es als Teil des Grußes annotiert.

Namen und Anonymisierung

- *Unterschrift*
 - ****|Unterschrift|****
 - Beispiele: ****|Gustav|**** oder ****|Erika [NN]|**** oder ****|Roland|**** mit ****|Elsa und Familie|****
 - Unterschrift beinhaltet jede Namensnennung der Absenderinnen und Absender mit oder ohne eigenständige Unterzeichnung
 - Die Unterschrift jeder Person wird einzeln getaggt. Unterschriften können nacheinander abgeschrieben und getaggt werden, ohne dass Absätze oder Zeilenumbrüche dabei beachtet werden müssen.
 - Unterschriftszusätze, die keinen Namen darstellen, wie *und Familie* gehören zu der direkt davor stehenden Unterschrift und werden nicht einzeln getaggt.
 - Possessivpronomen wie *Deine* und *Eure* gehören nicht zur Unterschrift, sondern zur Grußformel (vgl. Sektion Gruß).
- *Nachname*
 - Nachnamen außerhalb des Adressfelds werden durch [NN] ersetzt:
 - Beispiel 1 (in der Anrede): *Liebes Fräulein Müller* wird als **|Liebes Fräulein [NN]|** erfasst.
 - Beispiel 2 (in der Unterschrift nach dem Gruß): *Herzliche Grüße, Eure Familie Koller* wird als ****|Herzliche Grüße, Eure|**** ****|Familie [NN]|**** erfasst.
 - Nachnamen von öffentlichen Personen, z.B. Politikern, müssen nicht ersetzt werden.
- *Persönliche Informationen*
 - Informationen wie Mailadressen, Kontodaten, Telefonnummern etc. werden durch [vertraulich] ersetzt:
 - Beispiel: *Meine Telefonnummer hier ist 0123 456 789* wird als *Meine Telefonnummer hier ist [vertraulich]* erfasst.

Layout und Textstruktur

- *Zeilenumbruch*
 - Jede neue Zeile wird durch das einmalige Drücken der Eingabetaste markiert.
 - Beispiel: Abbildung 4 (a) wird als
 - Wir geniessen hier Sonne¶
 - Meer + Berge nach einem¶
 - eher nasskalten windigen¶
 - Ferienbeginn.¶erfasst
 - Zeilenumbrüche können auch leer gelassene Zeilen nach Absätzen markieren (vgl. Absatz, Abstand).
- *Absatz, Abstand*
 - Ein Absatz wird durch das zweimalige Drücken der Eingabetaste markiert.
 - Beispiel 1: Abbildung 5 als
 - **|Hay [Bettina]|** ¶¶
 - Hier in Scual ist¶

- traumhaftes Wetter.
- Am Mittwoch schlittelten
- wir auf Ftan. ¶

¶

- ***|Herzliche Schneegrüsse
- von|***| [Helene]|***

¶

- ***|Liebe Grüße an die ganze
- Familie|***

- o Beispiel 2: Abbildung 6 als

- **|Liebe [Leandra]|** ¶

¶

¶

- Von unserer wunder-
- schönen Reise

- o Beispiel 3: Abbildung 4 (b) als

- **|Salü zäme!|** ¶
- Wir geniessen hier Sonne
- Meer + Berge nach einem
- eher nasskalten windigen
- Ferienbeginn. ¶

¶

- ***|Frohe Weihnachten
- und es guets 2001

¶

- wünschen|***

- o Ein Absatz wird gesetzt, um folgende Merkmale zu markieren:

- größerer Abstand als sonst zwischen den Zeilen,
- eingerückte Passagen wie bspw. Anreden oder Datumsangaben,
- Zusätze zum Haupttext, die am Rand oder sonst irgendwo außerhalb des Haupttextes stehen.
- Faustregel: Alles, was sich visuell als Texteinheit abgrenzen lässt, wird als eigener Absatz markiert.
- Wenn mehrere Zeilen leer gelassen worden sind, wird ein Absatz gesetzt. Pro leere Zeile wird zusätzlich 1 Mal die Eingabetaste gedrückt (vgl. Zeilenumbruch).

- Unleserliches

- o [unclear]

- Beispiel: Die Statue ist wirklich [unclear].

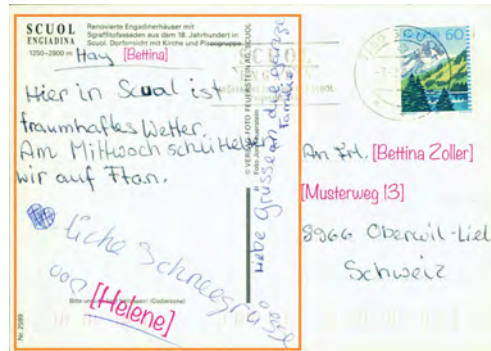
- o Dieser Tag umschließt nicht, sondern zeigt lediglich die Existenz einer unleserlichen Einheit an, die aus einem oder mehreren Wörtern bestehen kann.

Abb. 4: Zeilenumbruch, Absatz und Abstand



Quelle: [anko] 100003

Abb. 5: Absatz und Abstand



Quelle: [anko] 90015

Abb. 6: Absatz und Abstand



Quelle: [anko] 100061

3.3 Erfassung der Metadaten

Metadaten sind dokumentenbezogene Informationen, die unabhängig vom Abschreiben des Textkörpers erfasst werden. Für unterschiedliche Typen von Metadaten stehen in der Maske teilweise Vorgaben zum Auswählen, teilweise freie Textfelder zum Abschreiben zur Verfügung.

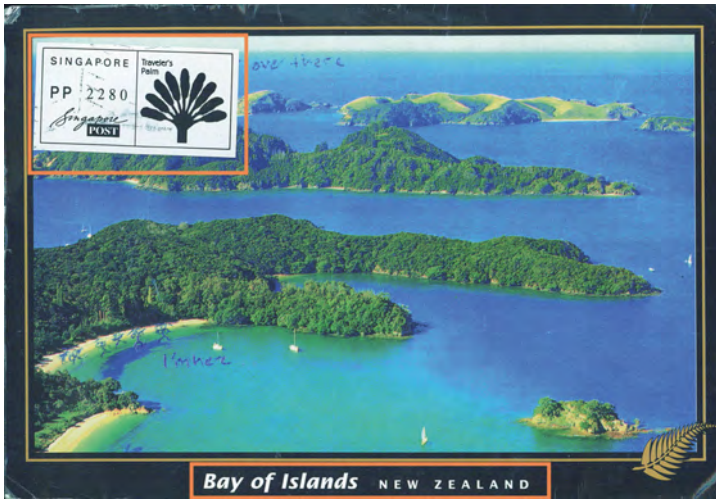
Anzahl/Geschlecht des Senders (Unterschriften)

- *Anzahl Sender*
 - auswählen: 1, 2, mehrere zählbare Individuen (z.B. Jan, Tom und Erika) oder unbekannt (z.B. Familie, Roland und Familie, WG, Büro)
 - Die Anzahl Sender entspricht der Anzahl an Unterschriften auf der Karte.
- *Geschlecht Sender*
 - auswählen: m (männlich), w (weiblich), gemischt (z.B. Jan und Erika), unbekannt (z.B. Roland und Familie, unbestimmte Vornamen wie Andrea oder Billy)
 - Das Geschlecht lässt sich i.d.R. aus den Unterschriften erkennen.

Gedruckter Text Bildseite und Ort/Land des Senders

- *Gedrucktes Bildseite*
 - exakt von links nach rechts und von oben nach unten abschreiben oder mit x versehen, wenn nicht vorhanden.
 - Aufkleber werden nicht abgeschrieben (vgl. Abbildung 7).
 - unleserliche Passagen werden durch [unclear] ersetzt.
 - Zeilenumbrüche (1 Mal Eingabetaste) und Absätze (2 Mal Eingabetaste) werden wie im Textkörper getippt. Verschiedene Texteinheiten werden durch Absatz voneinander getrennt.
 - Beispiel (vgl. Abbildung 7):
 - Bay of Islands¶
 - NEW ZEALAND
- *Ort Sender/Karte*
 - Stadt erfassen oder mit x versehen, wenn unbekannt
 - Beispiel: *Feldberg – Schwarzwald – Deutschland* als *Feldberg, Schwarzwald*
 - Der Ort des Senders lässt sich anhand des gedruckten Texts auf der Bild- oder der Schriftseite erschließen.
 - Wenn im gedruckten Text mehrere Orte genannt werden, sollen alle erfasst und mit Komma voneinander getrennt werden.
- *Land Sender/Karte*
 - Das Land des Senders wird aus der Drop-Down-Liste ausgewählt.
 - Das Land des Senders lässt sich anhand des gedruckten Texts auf der Bild- oder der Schriftseite erschließen. Sollte das Land nicht eindeutig ersichtlich sein, wird x ausgewählt.

Abb. 7: Aufkleber und Gedrucktes



Quelle: [anko] 70123

Verzierungen durch Sender

- *Verzierung Schriftseite*
 - o auswählen: vorhanden oder nicht vorhanden
 - o Hier werden alle vom Sender angebrachten Zeichnungen, Stempel, Ausstattungen, Sticker, Glitzerstaub o.Ä. erfasst.
- *Verzierung Bildseite*
 - o auswählen: vorhanden oder nicht vorhanden
 - o Hier werden alle vom Sender angebrachten Zeichnungen, Stempel, Ausstattungen, Sticker, Glitzerstaub o.Ä. erfasst.

Poststempel/Textkörper

- *Datum*
 - o Das Datum kann entweder aus dem Poststempel oder, wenn vorhanden, aus dem handschriftlichen Text übernommen werden. Format: in vorhandene Felder abschreiben (YYYY/MM/DD)
 - o Beispiel 1:
 - 1939
 - 07
 - 22
 - o Beispiel 2:
 - 1935
 - x
 - x
 - o Beispiel 3:
 - 197x

- 09
- 30
- o wenn Tag/Monat/Jahr nicht bekannt: Tag/Monat/Jahr mit *x* versehen
- o wenn einzelne Ziffern nicht lesbar sind, die einzelnen Ziffern mit *x* versehen

Adressfeld und Anzahl/Geschlecht des Empfängers

- *Anrede*
 - o Anrede exakt abschreiben oder mit *x* versehen, wenn nicht vorhanden
 - o Beispiele: Fräulein, An Fam., Frl., Frau Dr., Hr., An Señorita
 - o Die Anrede ist im Adressfeld ersichtlich und umfasst alles, was vor dem Eigennamen steht.
 - o Im Fall von *An Familie Meier* wird in dieses Feld nur *An Familie* geschrieben.
- *Empfänger*
 - o Vorname(n) und Nachname(n) exakt abschreiben. Alle Nachnamen werden in Großbuchstaben transliteriert; die Vornamen jeweils mit großem Anfangsbuchstaben.
 - o Beispiel 1: *Doris Müller* als *Doris MÜLLER*
 - o Beispiel 2: *HANS MÜLLER & F. KELLER* als *Hans MÜLLER & F. KELLER*
 - o Beispiel 3: *Klara + Anneli Huber* als *Klara + Anneli HUBER*
 - o Beispiel 4: *Otto Mayer GmbH* als *Otto MAYER GmbH*
 - o Die Namen müssen bei der Transkription nicht anonymisiert werden. Die erfassten persönlichen Daten werden in der Nachbearbeitung anonymisiert
- *Postleitzahl*
 - o Ziffern der Postleitzahl ohne Länderkürzel abschreiben oder mit *x* versehen, wenn unbekannt
 - o Beispiel: *CH-8902* wird als *8902* erfasst
 - o Die Postleitzahl ist im Adressfeld ersichtlich.
- *Ort Empfänger*
 - o Stadt/Ortschaft exakt abschreiben oder mit *x* versehen, wenn unbekannt
 - o Beispiele: *Zürich, Winterthur ZH, Löbau-Sachsen*
 - o Der Ort des Empfängers ist im Adressfeld ersichtlich.
 - o Zielorte werden, wenn möglich, auf Deutsch abgeschrieben. *Zurigo* wird als *Zürich*, *Munich* wird als *München* erfasst.
- *Land Empfänger*
 - o Das Land des Empfängers wird aus der Drop-Down-Liste ausgewählt.
 - o Das Land des Empfängers ist im Adressfeld ersichtlich.
- *Anzahl Empfänger*
 - o auswählen: 1, 2, mehrere zählbare Individuen oder unbekannt (z. B. Familie, WG, Büro etc.)
 - o Beispiel 1: *Otto, Lisa und Lea Mayer* als *mehrere zählbare Individuen* auswählen
 - o Beispiel 2: *Familie Otto Mayer* als *unbekannt* auswählen
 - o Die Anzahl Empfänger lässt sich aus der Anrede bzw. aus der Anzahl Namen im Adressfeld erschließen.

- *Geschlecht Empfänger*
 - o auswählen: m (männlich), w (weiblich), gemischt (z.B. Maria und Peter), unbekannt (z.B. Familie, WG, Büro etc.)
 - o Beispiel 1: *Andrea Rossi* als *unbekannt* auswählen, da Andrea sowohl ein Mann als auch eine Frau sein kann.
 - o Beispiel 2: *Familie Otto Mayer* als *unbekannt* auswählen
 - o Beispiel 3: *Otto, Lisa und Lea Mayer* als *gemischt* auswählen
 - o Das Geschlecht der Empfänger lässt sich aus der Anrede im Adressfeld oder den Vornamen erschließen.

3 Fazit

In diesem Anhang wurden die Richtlinien beschrieben für die Sortierung der Karten, das Abschreiben der Karten, die Annotation des Textkörpers und die Erfassung der Metadaten. Dieser Arbeitsprozess wurde durch das Schreibbüro auf der Grundlage der Online-Eingabemaske durchgeführt. Im Anschluss wurde die Qualität der eingegebenen Daten maschinell und manuell kontrolliert und verbessert. Das Total der erfassten digitalisierten Ansichtskarten liegt bei über 12.000 Karten. Sie bilden die Grundlage des linguistisch untersuchbaren Korpus. Aufbau und Nutzung des Ansichtskartenkorpus werden bei Sugisaki, Wiedmer, Calleri in diesem Band ausführlich dargestellt.

Literatur

Sugisaki, Kyoko/Wiedmer, Nicolas/Hausendorf, Heiko (2018): Building a Corpus from Handwritten Picture Postcards: Transcription, Annotation and Part-of-Speech Tagging. In: Nicoletta Calzolari et al. (Hg.): Proceeding of the 11th International Conference on Language Resources and Evaluation (LREC'18), 255–259.